DOI 10.18276/978-83-8419-028-9-36

THE PROBLEM OF DISCRIMINATION IN THE APPLICATION OF AI IN PUBLIC SERVICE MANAGEMENT

Kinga Flaga-Gieruszyńska[a]

[a]*University of Szczecin, Faculty of Law and Administration, Poland*

ABSTRACT

Purpose: *The aim of the study is to identify and analyse the mechanisms leading to discrimination as a result of the use of AI systems in public service management, as well as to identify possible solutions to reduce these phenomena.*

Need for the study: *With the rapid development of AI technologies, they are increasingly being used by public institutions to make decisions in areas such as welfare, education, security or employment. However, the lack of transparency of algorithms, biased input data and insufficient oversight mechanisms can lead to the reproduction and exacerbation of existing social inequalities, resulting in discrimination against specific groups of citizens.*

Methodology: *The analysis was based on a literature review, reports from public institutions, case studies and current examples of AI implementation in the public sector. Ethical and legal aspects were also taken into account, based on regulatory documents and recommendations on the use of AI in public administration.*

Findings: *The use of AI without proper oversight and awareness of risks can lead to the automation of discriminatory practices, especially towards ethnic minorities, people with low socio-economic status or people with disabilities. In practice, there is also often a lack of clear procedures regarding accountability for decisions made by algorithms and mechanisms for audit and correction.*

Practical Implications: *The results of the study point to the need to create a transparent and ethical regulatory framework, to ensure diversity in training data, and to implement algorithmic audits and accountability mechanisms. It also recommends public participation in the design and implementation of AI systems in public services to ensure fairness and equity.*

Keywords: AI, public service management, anti-discrimination, equality before the law, automated decision-making process

Jel codes: K23; K38; K41

1. INTRODUCTION

Recent years have witnessed the dynamic development of technologies based on artificial intelligence (AI), which are increasingly being applied in the public sector. Public administrations and public institutions are implementing algorithms to optimise decision-making processes, improve administrative efficiency and increase the availability of public services. AI-based systems are being used, for example,

Peer-review under responsibility of the scientific committee of 28th European Conference on Artificial Intelligence (ECAI 2025) and Intelligent Management Workshop.

to assess eligibility for social benefits, manage urban traffic, predict health needs, and for recruitment and selection in education systems. However, the implementation of such technologies is also associated with several risks, including the particularly serious problem of discrimination - both conscious and unconscious - against specific social groups.

A growing number of examples from different countries indicate that improperly applied algorithms can reinforce existing social inequalities rather than redress them. This is mainly due to imperfect training data, which reflects historical biases, and a lack of transparency in the operation of AI systems (the so-called 'black box'). The lack of clarified institutional accountability for erroneous algorithmic decisions and the limited possibility for a citizen to appeal a decision made by an AI system are also often a problem. As a result, artificial intelligence, which was supposed to serve as a tool to promote fairness and equality, can lead to systemic exclusion of those who are already socially and economically disadvantaged.

The main theses under scrutiny focus on three issues. First, AI systems used in the management of public services can reproduce and exacerbate existing social inequalities. Second, the lack of appropriate regulatory frameworks, oversight mechanisms and transparency of algorithmic decisions fosters the perpetuation of discriminatory practices. Third, both the design and implementation of AI in the public sector should be based on ethical principles, transparency and public participation.

The presentation uses as its main research method a critical analysis of the literature on the subject, including both academic publications in the social sciences and reports from public institutions. In addition, case studies illustrating concrete examples of AI applications in public institutions that led to discriminatory situations were included.

The paper is interdisciplinary and combines a technological and legal perspective with social analysis. The aim of the paper is not only to identify problems but also to formulate recommendations for the design and implementation of fair, inclusive and accountable AI systems in the public sector.

2. LITERATURE REVIEW

The application of AI in the public sector has been the subject of intense scrutiny in both the information technology and social science literature. Research indicates that while AI can bring significant benefits in terms of efficiency and precision to public institutions (Eubanks, 2018; OECD, 2021), it also carries the risk of reproducing social biases and inequalities (O'Neil, 2016; Noble, 2018).

One of the key problems discussed in the literature is the bias in training data. Authors such as Barocas and Selbst (2016) and Binns (2018) highlight that algorithms 'learn' from historical data, which often contain structural biases (e.g. race, class or gender), leading decision-making systems to replicate these inequalities. For example, Eubanks (2018) analyses cases from the US where AI systems used to assess social needs led to disproportionate surveillance of the poor and ethnic minorities. The literature also criticises the lack of transparency (the 'black box problem') and the difficulty of auditing algorithmic decisions. Pasquale (2015) points out that many AI systems used by public institutions operate in a non-transparent manner, making it difficult for citizens to understand decision-making mechanisms and to assert their rights in the event of erroneous or hurtful decisions.

In the context of public governance, there is also the issue of a lack of legal and ethical accountability for AI actions. Yeung (2018) and Veale and Brass (2019) highlight the legal gap regarding accountability for decisions made by algorithmic systems. They point to the need for regulatory frameworks and the implementation of ethical guidelines for the design of AI systems in public services. European literature (e.g. Floridi et al., 2018; European Commission, 2020) emphasises the need to implement AI ethical principles such as fairness, transparency, accountability and inclusivity. It is noted that the implementation of technologies should be done with the participation of citizens and experts from different fields, based on a participatory and intersectional approach.

There are also empirical studies analysing specific use cases of AI in the public sector. Among the most cited are the COMPAS system used in the US to predict the risk of recidivism (Angwin et al., 2016), Amazon's recruitment algorithm that excluded women (Dastin, 2018), and the SyRI system in

the Netherlands that automatically classified benefit recipients as potential fraudsters, leading to a court ruling declaring the practice a human rights violation (UN Special Rapporteur, 2020).

In conclusion, the literature unanimously points to the need for critical reflection and regulation of the use of AI in the public sector. A consensus is emerging that technological innovations alone do not guarantee social justice - on the contrary, in uncontrolled conditions, they can perpetuate and exacerbate discrimination. It is, therefore, necessary to develop interdisciplinary research that takes into account technological as well as social, legal and ethical aspects.

3. METHODOLOGY

This paper is based on a qualitative method to analyse in depth the phenomenon of discrimination related to the use of artificial intelligence in public service management. An interdisciplinary approach was used, combining the perspectives of the social sciences, legal sciences and information technology, with a particular focus on the ethics of technology and public management.

The primary research method was a critical review of the literature on the subject. Both academic texts - monographs, articles and peer-reviewed publications - and reports from public organisations were analysed. The selection of sources was purposive - the focus was on works that relate directly to the application of algorithmic systems in the context of public administration, with a particular focus on ethical and social aspects. It was also important to include publications critical of the unreflective implementation of AI, which point to its discriminatory and anti-social potential.

In the following part of the study, case studies were used as a complementary method. Known and well-documented examples of AI systems that have in the past led to practices considered discriminatory or violative of human rights were analysed. For detailed analysis, the case of the EPV-R (Evaluation of Partner Violence Risk) tool, implemented in the courts of the Basque Country (Spain) to support the operation in domestic violence cases, was used. This predictive tool supports the decisions of judges and law enforcement officers in domestic violence cases. A study by Valdivia at al. (2024) found that a lack of adequate user training and limited algorithm transparency led to mechanisms for automatic profiling of female victims and influenced decisions to apply protection and supervision measures. The second case was the SyRI system used in the Netherlands to type people suspected of defrauding social security benefits - its operation was found by the court to violate the right to privacy and the principles of non-discrimination. The third was Amazon's internal recruitment system, which, although developed in the private sector, revealed mechanisms of bias against women, a good illustration of the risk of historical bias being transferred to technology.

The research also analysed selected normative documents and regulatory guidelines on the use of artificial intelligence in public administration. Among others, the European Commission's AI White Paper, the OECD Principles for Responsible Artificial Intelligence and the guidelines developed by the European Commission's independent AI expert group were taken into account. The aim of this part of the analysis was to determine to what extent the existing ethical and legal frameworks respond to the challenges of algorithmic discrimination and to what extent they are applied in practice.

It is worth noting that due to the fact that the work covers issues in legal sciences, the study is exploratory and based on secondary sources only. No in-house empirical research or technical testing of algorithmic systems was carried out due to the dissimilarity of the analysis in the field of legal sciences. It is also limited by the fact that many AI systems used in administration operate in a non-transparent (black box) manner, which makes it difficult to fully evaluate them. Nevertheless, the collected material allows a thorough analysis of the phenomenon of discrimination and the formulation of conclusions and recommendations of a general nature, useful for further research and legislative work.

4. RESULTS

The results of the analyses conducted confirm that the use of AI in the management of public services can lead to unintentional but systemic forms of discrimination. Although AI is often seen as a neutral and objective tool, in reality, it embodies and reproduces the biases contained in the data on which it is trained. The case studies of EPV-R, SyRI and Amazon confirm that the lack of transparency of

algorithmic, historical data biases and limited end-user awareness can result in socially unjust and legally controversial decisions.

The EPV-R system (Evaluación del Riesgo de Violencia de Pareja - Revisada) was implemented in the civil judiciary of the Basque Country as a tool to support the assessment of domestic violence risk. Its aim was to support judges and law enforcement agencies in deciding whether to grant protection to women at risk of partner violence. However, research by Valdivia, Hyde-Vaamonde and García-Marcos (2024) found that the algorithm often classified serious cases as low risk - even in situations where violence later escalated to violence or homicide. The system was based on a limited set of characteristics, without taking into account relevant cultural and social factors, and its users were not trained to interpret the results (Valdivia et al., 2024). The process ran in parallel with the forensic investigation. The police had their own protocol for assigning protection measures to victims, even if the case had not yet been decided by a judge, and the case was handled by the police depending on the risk assessment assigned to each person. The assessment could change if the police managed to collect more information from the victim, the perpetrator or the neighbourhood, but the EPV-R was always involved in the process. The results of the questionnaire were, in all cases, included as evidence in the report received by the judges, which did not happen in other similar systems. This meant, in practice, that judges could face a dichotomy between their own opinion of the level of risk and that given by the algorithm.

It was also controversial that the EPV-R ranked risk higher for women from other cultures - indicating the algorithm's potentially biased approach to ethnicity. The system was accepted by judges as 'scientific' and reliable, limiting the possibility of critical evaluation in practice. The questions asked of victims of violence in the questionnaire are not necessarily asked in numerical order. The first question on the list is only worth one point but refers to whether the victims themselves or the perpetrators are 'foreigners' - except that the word used does not refer to all foreigners. In fact, this question is about people of a non-Western culture. According to the assumptions of the developers of this system, this would only apply to all cases where there is a non-European cultural understanding in relation to couples. Significantly, however, neither a specific methodology nor a list of referenced countries is used when completing the questionnaire and the selection of officers is subjective. (Bellio, 2023).

The SyRI system (Systeem Risico Indicatie), used in the Netherlands to detect abuses in the social benefit system, was officially banned in 2020 after a court ruling in The Hague. SyRI analysed residents' social and financial data in order to target people 'suspected' of defrauding benefits. In practice, however, the system focused almost exclusively on low socio-economic neighbourhoods with a high proportion of migrants, resulting in biased profiling of entire communities (Zuiderveen Borgesius, 2020). The court found that SyRI violated fundamental human rights - including the right to privacy and protection of personal data - and did not meet the standards of transparency and proportionality required by European law. This was one of the first cases in which the court directly invoked human rights principles to prohibit the use of an algorithm in the public sector (Hussain, 2020).

Another relevant example of the use of artificial intelligence in the management of public services, which has led to discriminatory situations, is the case of an automatic fraud detection system in the Swedish social security system. The state-run Försäkringskassan (Social Insurance Agency) started to use an algorithmic risk assessment model to identify individuals who might be taking unfair advantage of public benefits (Molén, 2024).

This system assigned a so-called 'risk score' to each applicant, based on a range of demographic and socio-economic data, such as migration status, education level, income level or age. Those with the highest scores were sent to a special screening department, which automatically triggered a screening process, often without the applicant's knowledge. These individuals were subjected to invasive vetting, involving, among other things, social media analysis, contact with educational institutions or banks, and even community interviews (Molén, 2024; Skelton, 2024).

As a result, there were situations in which people with low social status, women, people with so-called 'foreign backgrounds', and people with disabilities were significantly more likely to be flagged as 'risky' by the algorithm - despite the lack of objective reasons for such assessments. A journalistic investigation by Lighthouse Reports, conducted in collaboration with the daily Svenska Dagbladet, found that the system operated in a systemically biased way, reinforcing stereotypes and social inequalities (Molén, 2024).

Amnesty International, which investigated the case in a human rights context, unequivocally found the operation of the algorithm to be discriminatory and incompatible with international law. The organisation called on the Swedish authorities to immediately halt the system, pointing out that it violates fundamental rights, including the rights to equal treatment, privacy and social security (Amnesty International, 2024). Despite previous warnings, including from the Inspectorate for Social Security (ISF), which already in 2018 drew attention to the unequal impact of the algorithm, the system continued to be used without significant modifications (AIAAIC, 2025). From a regulatory perspective, it is worth highlighting that the system - used in a non-transparent manner and without audit mechanisms - may not be compliant with the new EU legal framework. According to the EU AI Act regulation, systems used in the public sector to profile citizens must be subject to strict requirements of transparency, risk assessment and protection against discrimination. The absence of these elements in the Swedish model makes its continued use highly problematic both ethically and legally (Babl.ai, 2024).

The case of Sweden, a country commonly associated with high levels of transparency and protection of citizens' rights, clearly demonstrates that even in the most advanced democracies, algorithms can operate in an oppressive manner if they are not properly designed, supervised and controlled. At the same time, it demonstrates the need to create independent auditing mechanisms and provide citizens with real opportunities to challenge decisions made by automated systems.

This kind of situation is also happening in the private sector. In 2018, it was revealed that Amazon was internally testing an algorithmic recruitment system to automate the assessment of job candidates. The problem was in the training data: the system learned from CVs sent to Amazon over the past 10 years, which reflected the company's dominant culture of hiring men in technical departments. As a result, the algorithm 'drew' the erroneous conclusion that men were a desirable profile and women a less desirable one (Vincent, 2018). In practice, this meant that the system downgraded CVs containing words such as 'women's' (e.g. 'women's chess club captain' or 'women's college'), while upgrading candidates who used technical language typical of male work environments, such as the terms 'executed', 'captured', 'driven' (Goodman, 2018). Furthermore, the tool ignored neutral competency data and favoured a particular style of expression that correlated with gender (Dastin, 2018; Vincent, 2018). Although the system was withdrawn before full implementation, this case perfectly demonstrates how easily artificial intelligence can transform existing social biases into automated, 'technical' discriminations. The 'garbage in, garbage out' mechanism was particularly evident here - the system not only mapped historical managerial preferences but systematised and reproduced them.

Each of the cases analysed highlights a different aspect of this phenomenon. The EPV-R system in the Basque Country shows how an inadequately implemented predictive tool can lead to a misjudgement of risk and expose victims of violence to further danger. The case of SyRI in the Netherlands reveals how the combination of big data and covert scoring can violate basic civil rights, leading to profiling and discrimination against entire communities. The Swedish model of welfare fraud detection, on the other hand, shows that even in countries with developed institutional mechanisms, AI can serve as a tool for systemic exclusion if it is not subject to independent oversight and transparency. The Amazon example points to the fundamental importance of input and shows how easily past patterns of inequality can be automated and made credible by an 'intelligent' system.

All these cases have a few key problems in common. The first is a lack of transparency and auditability, as the algorithms used in the public sector are often 'black boxes' whose operation is not understood by citizens or even by those implementing them. There is a lack of mechanisms for independent scrutiny, fairness and equity compliance tests (Binns, 2018; Veale & Brass, 2019). Also not insignificant is flawed or biased training data - the data used to train AI systems reflects existing social inequalities, historical institutional biases and lack of representation of marginalised groups (Barocas & Selbst, 2016; Noble, 2018). Consequently, even seemingly neutral algorithms can lead to systematic exclusion. This leads to the fundamental problem of a lack of institutional accountability - in the cases studied, there were no clear procedures in place to challenge the algorithm's decisions or assign responsibility for their consequences. As a result, affected citizens often had no real possibility to defend their rights. Moreover, the technocratisation of public decisions is an important threat. There is a tendency in public services to fetishise 'intelligent' systems, which are treated as more reliable than human judgment, even if they are based on incomplete data and models that do not fit the social context. Public administrations mistakenly treat the outcome of an algorithm as a 'neutral fact', undermining citizen scrutiny and democratic accountability of the administration (Pasquale, 2015; Eubanks, 2018).

From the above, it follows that technology in itself is not neutral. Designing and implementing AI systems in public administration requires not only technical competence but, above all, social and ethical awareness. Data-driven decision-making systems should be developed in a participatory manner, involving representatives of different social groups, human rights experts, as well as NGOs. Moreover, there is an urgent need for a clear and enforceable regulatory framework. The AI Act Regulation, adopted at the European Union level, is an important step in this direction, but its effectiveness will depend on the implementation mechanisms at the national level. It will be crucial not only to classify systems as high-risk, but also to ensure that they can be realistically controlled, suspended or banned in the event of a breach of citizens' rights.

5. DISCUSSION

The introduction of artificial intelligence into the public service sector has spawned a new quality in the functioning of public institutions - more automated, 'objectifying' and data-driven. However, at the same time, it has brought to light phenomena that have hitherto occurred mainly in social and institutional relations but have now been perpetuated and amplified by algorithmic decision-making mechanisms. Discrimination that might previously have been incidental becomes, thanks to AI, systematic, difficult to detect and - most importantly - difficult to challenge.

However, the discussion of the problem of algorithmic discrimination should not end with its identification. The key issue relates to what action should be taken to ensure that technology does not serve to exacerbate social inequalities, but to redress them. One of the most important lessons to emerge from the analysis of past cases is the need to change the approach to the design and implementation of technology in public institutions. Currently, many AI implementations in the public sector follow a logic of 'technocratic efficiency': the system is supposed to work faster, cheaper and without 'human error'. Meanwhile, many algorithmic errors are not due to accidental mistakes, but to a systemic lack of control, context and ethical imagination. In this sense, a paradigm shift is required - from human-dominant technology to human-serving technology. To achieve this, multi-level action is required. At the institutional level, this means the need to create sustainable procedures for assessing risks and social impacts even before AI systems are deployed. Instead of testing tools on citizens, public administrations should adopt a preemptive model, in which any decision to implement a new system is preceded by stakeholder consultations, legal and ethical analysis, and a forecast of possible impacts on different social groups.

The next step is to create structures of accountability - not only legal, but also organisational and political. It must be made clear who is responsible for the decision that the algorithm takes, as this strengthens citizens' sense of legal security. In this regard, it is also important that there is a possibility to appeal against this decision and that citizens are given insight into the mechanism that evaluates, assigns and classifies them. These issues are too often ignored or left unaddressed today. Meanwhile, every action of an AI system should have a 'face' - a specific person or organisational unit that is responsible for its effects and makes the decision to continue or stop it.

Public scrutiny of the technology can play an important role in this process. Independent audits of algorithms, the availability of technical documentation, citizen reporting of errors or whistleblower mechanisms can all build the resilience of democratic institutions to technological errors and abuse. At the same time, it helps counteract the loss of public trust, which can be exacerbated when administrative decisions become illegible, irrevocable and depersonalised.

Citizen participation in the system design process is also extremely important. Including representatives of community organisations, minority groups, academia and end-users in consultations can help identify weaknesses in a system before its operation leads to irreversible damage. In many countries, 'data democratisation' and 'participatory AI' initiatives are emerging, which involve the co-creation of technology policies with the participation of a broad spectrum of civil society. In this way, it becomes possible not only to increase the effectiveness of digital tools, but above all, their social legitimacy.

Finally, it is worth noting that technology alone will not solve problems that are deeply embedded in social structures. Even the best-designed algorithmic system can act discriminatorily if it is based on data from an unjust social order. This means that the fight against algorithmic discrimination cannot be

separated from the broader fight for social equality. AI cannot be treated as a neutral technology - it is a political tool that shapes the relationship between citizens and institutions, affects access to resources and participation in public life.

Therefore, corrective action must be carried out in parallel at several levels: technological, institutional, legal and cultural. We must learn to think of algorithms not just as technical products, but as social decisions encoded in the language of computer science. Only then will it be possible to create public systems that not only 'work' but also work fairly.

6. CONCLUSIONS

The analyses carried out clearly show that the use of artificial intelligence systems in the management of public services can lead to serious social consequences if these technologies are not properly supervised and embedded in a legal and ethical context. In each of the cases described - from benefit selection algorithms in Sweden to abuse detection tools in the Netherlands - the risk of replicating and reinforcing inequalities that previously operated as implicit patterns of social treatment became apparent.

Based on the data collected, the key problem is not the technology itself, but the way it is implemented. The main mistakes made by public institutions are the lack of transparency, insufficient control over data quality and the unavailability of viable redress mechanisms for citizens. As a result, decisions made by algorithms become virtually unassailable - violating the basic principles of procedural fairness and trust in the state.

To counter these phenomena, public administrations should deploy AI based on clear criteria: any technological solution should undergo a formal risk assessment, include built-in audit mechanisms and allow the system's operation to be explained to the individual concerned. Furthermore, key decisions - such as granting a benefit, classifying a citizen or interfering with his or her life situation - should not be made solely by a machine. It is necessary to ensure that a human retains ultimate control over the decision-making process ('human in the loop').

In light of planned and already implemented European regulations, such as the AI Act, Member States will be required to identify so-called high-risk systems, implement codes of conduct and conduct oversight of algorithmic systems used in the public sector. However, the experiences described in the study show that legislation is not enough - what is also needed is a change in institutional practices, greater involvement of ethics and human rights experts and, above all, the inclusion of citizens in the evaluation and consultation of technological solutions.

The most important conclusion of this analysis, then, is that technological justice will not emerge spontaneously because of AI development. It requires deliberate action - responsible design, informed management and consistent enforcement of equal treatment. Otherwise, rather than fostering the development of public services, AI may deepen social divisions and undermine trust in democratic institutions. The responsibility to counter these effects lies with technology designers as well as policy makers, regulators and civil society. What is at stake in this debate is not just the efficiency of digital government, but the quality of life of citizens and the shape of the relationship between the state and the individual in an era of digital transformation.

REFERENCES

AIAAIC. (2025, April). *Sweden fraud prediction algorithm found to discriminate against women, migrants*. Retrieved from: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents /sweden-fraud-prediction-algorithm-found-to-discriminate-against-women (15.04.2025).

Amnesty International. (2024, November). *Sweden: Authorities must discontinue discriminatory AI systems used by welfare agency*. Retrieved from: https://www.amnesty.org/en/latest/news/2024/11/sweden-authorities-must-discontinue-discriminatory-ai-systems-used-by-welfare-agency (15.04.2025).

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica.* Retrieved from: https://www. propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (12.04.2025).

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*(3), 671-732. https://doi.org/10.2139/ssrn.2477899.

Bellio, N. (2023). *How an algorithm to assess the risk of gender-based violence sees people from 'different cultures*'. AlgorithmWatch. Retrieved from: https://algorithmwatch.org/en/basque-country-gender-violence-algorithm (15.03.2025).

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency (FAT*), 149-159. Retieved from: https://proceedings.mlr.press/v81/binns18a/binns18a.pdf (15.03.2025).

Dastin, J. (2018). *Insight* - Amazon scraps secret AI recruiting tool that showed bias against women, *Reuters*. Retireved from: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (12.04.2025).

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. *St. Martin's Press*.

European Commission. (2020). *White Paper on Artificial Intelligence: A European approach to excellence and trust.* Retrieved from: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (12.04.2025).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines,* 28, 689-707. https://doi.org/10.1007/s11023-018-9482-5.

Goodman, R. (2018, October 12). *Why Amazon's automated hiring tool discriminated against women*. ACLU. Retrieved from: https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against (12.04.2025).

Hussain, H. (2020). *Dutch Court Finds SyRI Algorithm Violates Human Rights Norms in Landmark Case*. Retrieved from: https://www.humanrightspulse.com/mastercontentblog/dutch-court-finds-syri-algorithm-violates-human -rights-norms-in-landmark-case (12.04.2025).

Molén, T. (2024). Sweden's Suspicion Machine. *Lighthouse Reports*. Retrieved from: https://www.lighthousereports.com/investigation/swedens-suspicion-machine

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. *NYU Press.* https://doi.org/10.2307/j.ctt1pwt9w5 (12.04.2025).

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. *Crown Publishing Group.*

OECD. (2021). The OECD Framework for the Classification of AI Systems. *OECD Publishing.* Retrieved from: https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9ecaen.htm (12.04.2025).

Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. *Harvard University Press.*

Skelton, S.K., (2024, November 29). Swedish authorities urged to discontinue AI welfare system. *Computer Weekly.* Retrieved from: https://www.computerweekly.com/news/366616576/Swedish-authorities-urged-to-discontinue-AI-welfare-system (12.04.2025).

UN Special Rapporteur on Extreme Poverty and Human Rights. (2020). - Report of the Special Rapporteur on extreme poverty and human rights. *United Nations Human Rights Council.* Retrieved from: https://www.ohchr.org/en/documents/thematic-reports/a74493-digital-welfare-states-and-human-rights-report-special-rapporteur (12.04.2025).

Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. *Algorithmic Regulation (edited by Karen Yeung and Martin Lodge), Oxford University Press*. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3375391 (12.04.2025).

Valdivia, A., Hyde-Vaamonde, C., & García Marcos, J. (2025). Judging the algorithm: A case study on the risk assessment tool for intimate partner violence implemented in the Basque Country. *AI & Society, 40*, 2633-2650. https://doi.org/10.1007/s00146-024-02016-9.

Werner, J. (2024). Swedish welfare agency's AI system faces accusations of discrimination and bias. Retrieved from: https://babl.ai/swedish-welfare-agencys-ai-system-faces-accusations-of-discrimination-and-bias (15.04.2025).

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4), 505-523. https://doi.org/10.1111/rego.12158.

Zuiderveen Borgesius, F. van. (2020). Digital welfare fraud detection and the Dutch SyRI judgment. *IAPP News / Human Rights Pulse.* Retrieved from: https://iapp.org/news/a/digital-welfare-fraud-detection-and-the-dutch-syri-judgment (12.04.2025).